Working Paper No. 138

CONSTRAINED MAXIMUM LIKELIHOOD ESTIMATION OF N STOCHASTICALLY ORDERED DISTRIBUTIONS

by

ARTHUR M. GEOFFRION

July, 1968

OPERATIONS RESEARCH DIVISION
WESTERN MANAGEMENT SCIENCE INSTITUTE
University of California, Los Angeles

41

University of California

Los Angeles

Western Management Science Institute

Working Paper No. 138

"CONSTRAINED MAXIMUM LIKELIHOOD ESTIMATION OF
N STOCHASTICALLY ORDERED DISTRIBUTIONS"

by

Arthur M. Geoffrion

July, 1968

This work was supported by the United States Air Force under Project RAND, and was carried out under a consulting arrangement with The RAND Corporation.

SUMMARY

This study considers the problem of determining step function maximum likelihood estimates for N stochastically ordered distributions, subject to the constraint that the estimates themselves must also be stochastically ordered. This problem arises, for example, in the context of reliability growth. Brunk, et al., [2] has achieved a closed form solution for the case $N = 2$, but was unable to extend the results to the case $N > 2$. We shall present a new analytical method based on the Kuhn-Tucker optimality conditions for the equivalent concave program. For $N = 2$, the method yields the closed form solution of Brunk, and for $N \geq 3$, the method yields--when used in conjunction with a reduction strategy developed in [4]--an efficient computational algorithm. The algorithm involves solving a short sequence of essentially unconstrained sub-problems with many fewer variables, and has been implemented and tested extensively on the IBM-7044. Computational experience is presented showing that large problems can be solved in reasonable time with good accuracy, especially when compared with the performance of a general nonlinear programming algorithm applied directly to the equivalent concave program. These results should also interest those concerned with solving large structured nonlinear programs (the largest we have solved involves 381 variables and 123 linear constraints), since the reduction strategy used here is of quite general applicability.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the contributions of E. M. Scheuer, who stimulated the author's interest in the present problem, and of S. P. Azen, who successfully implemented the computational approach suggested here as the ESOD-3 program.

CONTENTS

I. INTRODUCTION

Let the random variable X^j have a (right-continuous) cumulative distribution function F^j, $j = 1,\dots,N$, independent and stochastically ordered* according to $F^1 \geq F^2 \geq \dots \geq F^N$. Maximum likelihood estimates $\hat{F}^j$ of the distributions are desired, based on the independent observations X_i^j ($j = 1,\dots,N$; $i = 1,\dots,n^j$) subject to the constraint $\hat{F}^1 \geq \dots \geq \hat{F}^N$.

When $N = 2$, explicit formulae for $\hat{F}^1$ and $\hat{F}^2$ were derived by Brunk, et al. [2]. They were unable to extend their results to the case $N \geq 3$. This study makes use of the Kuhn-Tucker Theorem of concave programming [3] to furnish an alternate derivation of the results in [2], and to derive a partial characterization of the solution for the case $N \geq 3$. We then show how the author's reduction strategy [4,5] can be used to take advantage of the analysis to achieve an effective computational solution for $N \geq 3$ and large numbers of observations. The computational procedure reduces the constrained problem to a short sequence of essentially unconstrained sub-problems with many fewer variables.

We begin by reformulating the given problem as one of concave programming with a finite number of variables and constraints. The analysis of the resulting concave program, with the help of the Kuhn-Tucker Theorem, is carried out in Sec. III. In Sec. IV we describe

---

*We write "$F \geq G$" for two distributions if $F(x) \geq G(x)$ for all x.

a computational method based on the reduction procedure. The method has been implemented and tested, and in Sec. V we give computational evidence that it is very effective in terms of computing time, storage requirements, and numerical accuracy. Details regarding the program, ESOD-3, are available in [1].

II. TRANSFORMATION OF THE PROBLEM INTO A CONCAVE PROGRAM

The constrained maximum likelihood estimate $\hat{F}^j$ must be sought from the class of discrete distributions, or the likelihood product has no upper bound. Let f(x) be the probability assigned to the value x by a discrete distribution F(x). Then we must maximize the likelihood product

$$\prod_{j=1}^{N} \prod_{i=1}^{n^j} f^j(X_i^j)$$

over all discrete distribution functions with the desired stochastic ordering. The task of this section is to reduce this problem to one with a finite number of variables and constraints.

For $j = 1,\ldots,N$, define the following: $x_1^j < x_2^j < \ldots < x_{n_j}^j$ to be the distinct values taken on by the X_i^j; x_o^j to be the smallest of the X_i^J for any J such that $j \le J \le N$; and $x_{n_j+1}^j$ to be the largest of the X_i^J for any J such that $1 \le J \le j$. We assert that it suffices to seek the maximum of the likelihood product over all discrete distributions having the desired stochastic ordering and also having the additional property $\mathcal{P}$.

<u>Property $\mathcal{P}$</u>: For $j = 1,\ldots,N$, $f^j(x) = 0$ unless $x \in \{x_o^j, x_1^j,\ldots,x_{n_j+1}^j\}$. This assertion follows from lemma 1.

<u>Lemma 1</u>: If $F_o^1,\ldots,F_o^N$ are any discrete distributions stochastically ordered as $F_o^1 \ge \ldots \ge F_o^N$, then there exist discrete distributions $F_*^1,\ldots,F_*^N$ stochastically ordered as $F_*^1 \ge \ldots \ge F_*^N$ with property $\mathcal{P}$ and at least as large a value of the above likelihood product.

<u>Proof</u>: We shall generalize an argument used in [2]. Define

$$F_*^1(x) = \begin{cases} 0 & , \text{ if } x < x_o^1 \\ F_o(x_{i+1}^1-), & \text{if } x_i^1 \le x < x_{i+1}^1,\ i = 0, 1, \ldots, n_1-1 \\ 1 & , \text{ if } x \ge x_{n_1}^1 \end{cases}$$

and

$$F_*^j(x) = \begin{cases} 0 & , \text{ if } x < x_o^j \\ \operatorname{Min}\{F_o^j(x_{i+1}^j-),\ F_*^{j-1}(x_i^j)\} & , \text{ if } x_i^j \le x < x_{i+1}^j,\ i = 0, 1, \ldots, n_j \\ 1 & , \text{ if } x \ge x_{n_j+1}^j \end{cases}$$

for $j = 2,\ldots,N$, where it is understood that any vacuous intervals in the definition of the F_*^j are to be ignored (e.g., if $x_o^2 = x_1^2$, then $[x_o^2, x_1^2)$ is empty). Clearly the F_*^j have property $\mathcal{P}$ and the desired stochastic ordering. To see that the F_*^j are true (increasing) distribution functions and that they have at least as large a likelihood product as the F_o^j, it suffices to show $f_*^j(x_i^j) \ge f_o^j(x_i^j)$ for $i = 1,\ldots,n_j$ and $j = 1,\ldots,N$. This is clearly true for $j = 1$. Consider now $2 \le j < N$. Note that $F_*^1 \ge F_o^1$ by construction, which implies $F_*^2 \ge F_o^2$, which implies $F_*^3 \ge F_o^3$, ..., which implies $F_*^N \ge F_o^N$. Thus for $i = 0, 1,\ldots,n_j$, we have

$$F_o^j(x_{i+1}^j-) \ge \operatorname{Min}\left\{ F_o^j(x_{i+1}^j-),\ F_*^{j-1}(x_i^j)\right\} \ge F_o^j(x_i^j),$$

since

$$F_*^{j-1}(x_i^j) \ge F_o^{j-1}(x_i^j) \ge F_o^j(x_i^j) \quad \text{and} \quad F_o^j(x_{i+1}^j-) \ge F_o^j(x_i^j).$$

It follows that for $i = 1,\ldots,n_j$,

$$f_*^j(x_i^j) = F_*^j(x_i^j) - F_*^j(x_i^j-) \ge F_o^j(x_i^j) - F_o^j(x_i^j-) = f_o^j(x_i^j).$$

This completes the proof.

Thus we have reduced the problem to one with less than $\sum_{j=1}^{N}(n_j+2)$ variables: $f^j(x_o^j), f^j(x_1^j), \ldots, f^j(x_{n_j+1}^j)(j = 1,\ldots,N)$. We say less than because some of the x_o^j coincide with x_1^j, or $x_{n_j+1}^j$ with $x_{n_j}^j$ (e.g., it is obvious that $x_{n_1+1}^1 = x_{n_1}^1$ and $x_o^N = x_1^N$). For notational convenience we shall adopt the convention that the redundant end numbers among $x_o^j \le x_1^j < x_2^j < \ldots < x_{n_j}^j \le x_{n_j+1}^j$ have been eliminated and that n_j has been redefined where necessary so that the remaining x_i^j can be relabeled $x_1^j < \ldots < x_{n_j}^j$. The corresponding $f^j(x_i^j)$ are denoted by $y_1^j,\ldots,y_{n_j}^j$ $(j = 1,\ldots,N)$.

The constraints are of two varieties. The first requires the y_i^j to correspond to discrete probability distributions; all y_i^j must be nonnegative, and

$$\sum_{i=1}^{n_j} y_i^j = 1 \tag{1.j}$$

must hold for $j = 1,\ldots,N$. The second variety assures stochastic ordering:*

$$\sum_{i=1}^{k} y_i^j \ge \left[\sum_i y_i^{j+1} : x_i^{j+1} < x_{k+1}^j\right], \quad k = 1,\ldots,n_j-1, \tag{2.j}$$

for $j = 1,\ldots,N-1$. (For the sake of a linear notation, complicated restrictions of the domain of summation follow a colon immediately following the summand.)

---

*(2.j) can be interpreted as enforcing $F^j \ge F^{j+1}$ by requiring that

$$F^j(x_{k+1}^j-) \ge F^{j+1}(x_{k+1}^j-), \quad k = 1,\ldots, n_j-1.$$

A similar interpretation can be offered for (3.j).

An alternative to (2.j) is

$$(3.j) \qquad \left[\sum_i y_i^j : x_i^j \le x_k^{j+1}\right] \ge \sum_{i=1}^{k} y_i^{j+1}, \; k = 1,\ldots,n_{j+1}-1.$$

For the sake of being specific we shall use (2.j) in the sequel.

Letting h_i^j be the number of X^j observations with value x_i^j (h_1^j or $h_{n_j}^j$ may be 0) and using log likelihood, we have finally reduced our original problem to the concave program:

$$(P) \qquad \underset{y_i^j \ge 0}{\text{Maximize}} \sum_{j=1}^{N} \sum_{i=1}^{n_j} h_i^j \ln y_i^j \quad \text{subject to } (1.j),\ j = 1,\ldots,N \text{ and } (2.j),\ j = 1,\ldots,N-1.$$

A number of constraints among (2.j), $j = 1,\ldots,N-1$, are likely to be manifestly redundant, and can therefore be eliminated;* for each j, denote the k-indices of the remaining nonredundant constraints by $\mathcal{K}^j$. Moreover, certain of these nonredundant constraints can actually be written as equalities, as we now show.

<u>Lemma 2:</u> For $j = 1,\ldots,N$,

A. $h_1^j = 0$ implies $1 \in \mathcal{K}^j$, and the corresponding constraint is satisfied with equality in any optimal solution of (P).

B. $h_{n_j}^j = 0$ implies that the right-hand side of the last constraint in $\mathcal{K}^{j-1}$ is $(y_1^j + \ldots + y_{n_j-1}^j)$ and that this constraint is satisfied with equality in any optimal solution of (P).

<u>Proof</u>: By construction, h_1^j can vanish only for $j \in \{1,\ldots,N-1\}$, and when it does, $x_2^j > x_1^{j+1}$ necessarily holds. This implies that the first constraint of (2.j) is $y_1^j \ge y_1^{j+1} + \ldots$, where we have not

*It is a trivial matter, due to the very special structure, to identify the redundant constraints.

specified the right-hand side fully. This constraint is clearly not redundant, so $1 \in \mathcal{N}^j$. Suppose now that this constraint is amply satisfied at an optimal solution of (P). Then y_1^j can be decreased by a positive amount and y_2^j increased by the same amount without destroying feasibility in (P), as one can easily verify. But this improves the value of the loglikelihood function, which is impossible. This proves part A.

By construction, $h_{n_j}^j$ can vanish only for $j \in \{2,\ldots,N\}$, and when it does, $x_{n_j-1}^j < x_{n_{j-1}}^{j-1}$ necessarily holds. This implies that the last nonredundant constraint of $\mathcal{N}^{j-1}$ is $y_1^{j-1} + \ldots \geq y_1^j + \ldots + y_{n_j-1}^j$, where we have not specified the left-hand side fully. Suppose now that this constraint is amply satisfied at an optimal solution of (P). Then $y_{n_j-1}^j$ can be increased by a positive amount and $y_{n_j}^j$ decreased by the same amount without destroying feasibility in (P), as one can easily verify. But this improves the value of the loglikelihood function, which is impossible. This proves part B.

For $j = 1,\ldots, N-1$, let $\underline{\mathcal{N}}^j$ denote the constraints of $\mathcal{N}^j$ identified in Lemma 2. The results of this section are summarized and extended in Theorem 1.

<u>Theorem 1</u>: The original constrained maximum likelihood problem can be formulated as the concave program (P), which has a unique optimal solution $\{\hat{y}_i^j\}$. Each component of this solution is positive, and it satisfies the constraints in $\underline{\mathcal{N}}^j$ $(j = 1,\ldots,N-1)$ with strict equality.

<u>Proof</u>: The only assertion unproved thus far is that (P) has a unique optimal solution and that this solution has all positive components.

It is easy to construct a feasible solution of (P) with each component positive. For example, it is readily verified that the

following construction is possible: let $y_i^1 = 1/n_1$ $(i = 1,\ldots,n_1)$; let y_i^2 $(i = 1,\ldots,n_2)$ have any positive values summing to 1 consistent with the satisfaction of the constraints in η^1 considered as equalities;...; let y_i^N $(i = 1,\ldots,n_N)$ have any positive values summing to 1 consistent with satisfaction of the constraints in η^{N-1} considered as equalities. Let v be the value of the objective function of (P) for some positive feasible solution. Clearly $v > -\infty$. Thus we may impose the additional constraint $\sum_{j=1}^{N} \sum_{i=1}^{n_j} h_i^j \ln y_i^j \geq v(> -\infty)$ on (P). Since (1.j) places an upper bound of 1 on each y_i^j and thereby an upper bound of 0 on $\ln y_i^j$, this constraint requires that $h_i^j \ln y_i^j \geq v(> -\infty)$ for all i and j. For $h_i^j > 0$, this requirement is equivalent to $y_i^j \geq e^{(v/h_i^j)}(> 0)$. From Lemma 2 and constraints (1.j) it follows that y_i^j is bounded strictly away from 0 even when $h_i^j = 0$ (in this case the above requirement is vacuous, since $v < 0$). All this is to say that the y_i^j may be taken to be bounded strictly away from 0 in (P). By the continuity of the maximand on the compact thus-restricted feasible region, (P) is seen to have a positive optimal solution $\{\hat{y}_i^j\}$. From the strict concavity of $\ln y_i^j$ we conclude that $\hat{y}_i^j$ is unique when $h_i^j > 0$. This implies, with the help of Lemma 2, that $\hat{y}_i^j$ such that $h_i^j = 0$ is also unique. This concludes the proof.

We remark that (P) is amenable to solution by various existing concave programming algorithms. For example, the algorithms of Graves [6] and Rosen [7] seem appropriate. So does a family of approximate algorithms of the "separable programming" variety, in view of the linear separability of the objective function. (See, e.g., Miller [8].)

However, these algorithms do not take full advantage of the important special structure enjoyed by (P).

III. APPLICATION OF THE KUHN-TUCKER THEOREM

In this section we apply the Kuhn-Tucker Theorem [3] to (P) to recover the analytical solution of Brunk, et al. [2] for the case $N = 2$. We shall see that the method becomes unwieldy for the case $N \geq 3$, but nevertheless does yield analytical results leading to an effective computational procedure.

The Kuhn-Tucker Theorem can be applied to (P), in view of the results of the last section, to yield the following characterization of $\{\hat{y}_i^j\}$.

Theorem 2: A trial solution $\{y_i^j\}$ is an optimal solution of (P) if and only if all $y_i^j > 0$, and for some collection $S^j (j = 1,\ldots,N-1)$ satisfying $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$ there exist generalized Lagrange multipliers $\lambda_k^j (k \in S^j;\ j = 1,\ldots,N-1)$ and $\mu^j (j = 1,\ldots,N)$ such that $(S^j;\ y_i^j;\ \lambda_k^j,\ \mu^j)$ satisfies the corresponding Kuhn-Tucker conditions:

$$\text{(KT-1)}\quad \frac{\partial}{\partial y_i^j}\left(\sum_{j=1}^{N} \sum_{i=1}^{n_j} h_i^j \ln y_i^j + \sum_{j=1}^{N} \mu^j \left(\sum_{i=1}^{n_j} y_i^j - 1 \right) \right.$$

$$\left. + \sum_{j=1}^{N-1} \sum_{k \in S^j} \lambda_k^j \left[\sum_{i=1}^{k} y_i^j - \sum_{i} y_i^{j+1} :\ x_i^{j+1} < x_{k+1}^j \right] \right) = 0 \text{ for all } i \text{ and } j,$$

$$\text{(KT-2)}\quad \begin{cases} \displaystyle\sum_{i=1}^{n_j} y_i^j = 1, \quad j = 1,\ldots,N \\[2ex] \displaystyle\sum_{i=1}^{k} y_i^j = \sum_i y_i^{j+1} : x_i^{j+1} < x_{k+1}^j, \quad k \in S^j, \quad j = 1,\ldots,N-1, \end{cases}$$

$$\text{(KT-3)}\quad \sum_{i=1}^{k} y_i^j \geq \sum_i y_i^{j+1} : x_i^{j+1} < x_{k+1}^j, \quad k \in \mathcal{N}^j - S^j, \quad j = 1,\ldots,N-1,$$

$$\text{(KT-4)}\quad \lambda_k^j \geq 0, \quad k \in S^j - \underline{\mathcal{N}}^j, \quad j = 1,\ldots,N-1.$$

Our general strategy will be to determine from (KT-1) and (KT-2), for S^j satisfying $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$, what $\{y_i^j\}$ and $\{\lambda_k^j\}$ must be. For (KT-3) and (KT-4) to be satisfied, it will follow that the S^j must satisfy certain necessary and sufficient conditions. With the S^j so determined, it remains only to compute $\{\hat{y}_i^j\}$ from (KT-1) and (KT-2).

This strategy will be completely successful only for the case $N = 2$. We assume henceforth that $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$ $(j = 1,\ldots,N-1)$.

3.1 THE CASE N = 2

We shall simplify the notation somewhat in this subsection by writing S for S^1, λ_k for λ_k^1, and $\mathcal{N}$ for $\mathcal{N}^1$.

When the differentiation is performed, (KT-1) becomes

$$\text{(KT-1)}_i^1\quad \left[(h_i^1/y_i^1) + \mu^1 + \sum_{k\in S} \lambda_k : k \geq i\right] = 0, \qquad i = 1,\ldots,n_1,$$

$$(KT\text{-}1)_i^2 \quad (h_i^2/y_i^2) + \mu^2 - \sum_{k\in S} \lambda_k : x_i^2 < x_{k+1}^1 = 0, \quad i = 1,\ldots,n_2.$$

Let S be expressed as $\{k_1,\ldots,k_s\}$, where $k_1 < \ldots < k_s$, when S is not empty. We see from (KT-1) that when S is empty, the ratios (h_i^j/y_i^j) are constant for $i = 1,\ldots,n_j$ and $j = 1$ and 2. When S is not empty, (h_j^1/y_j^1) remains constant over the integer intervals $[1, k_1]$ $(k_1, k_2],\ldots,(k_s, n_1]$, while (h_j^2/y_j^2) remains constant over the integer intervals $\{j : x_j^2 < x_{k_1+1}^1\}$, $\{j : x_{k_1+1}^1 < x_j^2 < x_{k_2+1}^1\},\ldots,$ $\{j : x_{k_{s-1}+1}^1 \le x_j^2 < x_{k_s+1}^1\}$, $\{j : x_{k_s+1}^1 \le x_j^2\}$. Thus S induces a partition P of $\{1,\ldots,n_1\}$ and a partition Q of $\{1,\ldots,n_2\}$. We denote a generic cell of P by p, and the i^{th} cell by $p_i (i = 1,\ldots,s+1)$; similarly, we use the notations q and q_i for Q. Since there is a natural 1:1 correspondence between the cells of P and Q, we sometimes denote the cell of Q corresponding to cell p of P by q(p).

Thus to determine $\{y_i^j\}$ satisfying (KT-1), it suffices to determine the 2(s+1) ratios (h_i^j/y_i^j) corresponding to the cells of P and Q. Using an obvious notation, we denote these ratios by $(h/y)_p^1$ and $(h/y)_q^2$.

<u>Lemma 3</u>: If (KT-1) has a solution $\{y_i^j\}$, then for each $p \in P$, $(h_i^1/y_i^1) = (h/y)_p^1$ for all $i \in p$, and for each $q \in Q$, $(h_i^2/y_i^2) = (h/y)_q^2$ for all $i \in q$.

Let us turn now to (KT-2). With our new notation, we may rewrite (KT-2) as (KT-2A) and $(KT\text{-}2B)_p$ for each $p \in P$:

$$\text{(KT-2A)} \qquad \sum_{i=1}^{n_1} y_i^1 = 1,$$

$$(\text{KT-2B})_p \qquad \sum_{i \in p} y_i^1 = \sum_{i \in q(p)} y_i^2 \ .$$

By Lemma 3, $(KT\text{-}2B)_p$ can be written

$$(\text{KT-2B})_p \qquad (y/h)_p^1 \sum_{i \in p} h_i^1 = (y/h)_{q(p)}^2 \sum_{i \in q(p)} h_i^2 ,$$

where it is understood that if $h_1^1 = 0$, then

$$(y/h)_{p_1}^1 \sum_{i \in p_1} h_i^1,$$

which is undefined, is defined as simply y_1^1; and, similarly, that if $h_{n_2}^2 = 0$, then

$$(y/h)_{q_{s+1}}^2 \sum_{i \in q_{s+1}} h_i^2$$

is defined as simply $y_{n_2}^2$.

Thus far we have reduced (KT-1) and (KT-2) to $[2(s+1) + (s+1) + 1]$ independent equations (namely, $(KT\text{-}1)_k^1$ and $(KT\text{-}1)_k^2$ for $k \in S$, $(KT\text{-}1)_{n_1}^1$, $(KT\text{-}1)_{n_2}^2$, $(KT\text{-}2B)_p$ for $p \in P$ and (KT-2A)) in $[3(s+1) + s + 2]$ unknowns (namely, y_i^1 (respectively y_i^2) for one i in each cell of P (respectively Q), λ^k for $k \in S$, and μ^1, μ^2). We shall see that we can obtain a complete solution.

Consider the case $S = \phi$. We may assume, because of Lemma 2, that all $h_i^j > 0$. The equations (KT-1) and (KT-2) reduce to:

$$(h_{n_1}^1 / y_{n_1}^1) + \mu^1 = 0,$$

$$(h_{n_2}^2 / y_{n_2}^2) + \mu^2 = 0,$$

$$(y_{n_1}^1 / h_{n_1}^1) \sum_{i=1}^{n_1} h_i^1 = (y_{n_2}^2 / h_{n_2}^2) \sum_{i=1}^{n_2} h_i^2,$$

$$(y_{n_1}^1 / h_{n_1}^1) \sum_{i=1}^{n_1} h_i^1 = 1 .$$

Their solution is obviously

$$y_{n_j}^j = (h_{n_j}^j / \sum_{i=1}^{n_j} h_i^j), \qquad j = 1, 2, \tag{4.1}$$

$$\mu^j = - \sum_{i=1}^{n_j} h_i^j, \qquad j = 1, 2.$$

From $(KT\text{-}1)^1$, we finally have

$$y_i^1 = (h_i^1 / \sum_{j=1}^{n_1} h_j^1), \qquad i = 1, \ldots, n_1; \tag{4.2}$$

and from $(KT\text{-}1)^2$,

$$y_i^2 = (h_i^2 / \sum_{j=1}^{n_2} h_j^2), \qquad i = 1, \ldots, n_2. \tag{4.3}$$

It remains to consider the case $S \neq \phi$. From $(KT\text{-}1)^j_{n_j}$, we have for $j = 1, 2$,

$$(5)^j \qquad \mu^j = -(h^j_{n_j}/y^j_{n_j}).$$

Upon subtracting $(KT\text{-}1)^1_{k_i+1}$ from $(KT\text{-}1)^1_{k_i}$, one obtains

$$(6) \qquad \lambda_{k_i} = (h^1_{k_i+1}/y^1_{k_i+1}) - (h^1_{k_i}/y^1_{k_i}), \quad i = 1,\ldots,s.$$

From (KT-2B), (5), and (6) we see that we can easily express μ^1, μ^2, $\lambda_k (k \in S)$, $(y/h)^2_{q_i}$ $(i = 1,\ldots,s)$ and $y^2_{n_2}$ in terms of y^1_k $(k \in S)$ and $y^1_{n_1}$. The problem of obtaining a complete analytic solution to (KT-1) and (KT-2) thus reduces to solving for the last mentioned quantities.

For each $p \in P$, add $(KT\text{-}1)^1_i$ and $(KT\text{-}1)^2_j$ for some $i \in p$ and $j \in q(p)$. The result is

$$(7) \qquad (h/y)^1_p + \mu^1 + (h/y)^2_{q(p)} + \mu^2 = 0, \quad p \in P.$$

Substituting (KT-2B) and (5) into (7), one obtains for each $p \in P$

$$(8)_p \qquad (y/h)^1_p = \frac{\left(\sum_{i \in p} h^1_i + \sum_{i \in q(p)} h^2_i\right)}{\left[(h^1_{n_1}/y^1_{n_1}) + (h^2_{n_2}/y^2_{n_2})\right]\left(\sum_{i \in p} h^1_i\right)},$$

where it is understood that if $h^1_1 = 0$, then $(8)_{p_1}$ is replaced by

$$y^1_1 = \sum_{i \in q_1} h^2_i \Big/ \left[(h^1_{n_1}/y^1_{n_1}) + (h^2_{n_2}/y^2_{n_2})\right].$$

Equation $(KT\text{-}2B)_{p_{s+1}}$ can be used to eliminate $y^2_{n_2}$ from (8). The resulting expressions for y^1_k ($k \in S$), in terms of $y^1_{n_1}$, can be used with (KT-2A) to solve for $y^1_{n_1}$. The result is

$$y^1_{n_1} = [(\eta^1_{s+1} + \eta^2_{s+1})/\eta^1_{s+1}(\eta^1 + \eta^2)]\, h^1_{n_1}, \tag{9}$$

where

$$\eta^1_j = \sum_{i \in p_j} h^1_i \quad (j = 1,\ldots,s+1),$$

$$\eta^2_j = \sum_{i \in q_j} h^2_i \quad (j = 1,\ldots,s+1), \text{ and } \eta^j = \sum_{i=1}^{s+1} \eta^j_i \quad (j = 1,2).$$

Finally, we have from (8) and (9) for $i = 1,\ldots,s+1$

$$(y/h)^1_{p_i} = \frac{(\eta^1_i + \eta^2_i)}{(\eta^1 + \eta^2)\eta^1_i}, \tag{10_i}$$

where it is understood that if $h^1_1 = 0$, then $(10)_1$ is replaced by $y^1_1 = \eta^2_1/(\eta^1 + \eta^2)$. This is the desired analytic solution. For completeness, we record (using (10) in (KT-2B)) that for $i = 1,\ldots,s+1$

$$(y/h)^2_{q_i} = (\eta^1_i + \eta^2_i)/(\eta^1 + \eta^2)\eta^2_i, \tag{11_i}$$

where it is understood that if $h^2_{n_2} = 0$, then $(11)_{s+1}$ is replaced by $y^2_{n_2} = \eta^1_{s+1}/(\eta^1 + \eta^2)$.

Let us summarize our results as

Lemma 4: Let S satisfy $\underline{\mathcal{N}} \subseteq S \subseteq \mathcal{N}$. Then (KT-1) and (KT-2) have a unique solution $(y_i^j; \lambda_k; \mu^j)$ given by (10) and (11), (6) and (5). ((4) is subsumed by (10), (11) and (5) in the event that $S = \emptyset$.) Moreover, all $y_i^j > 0$ in this solution.

Now that we have an analytical solution of (KT-1) and (KT-2) for any S satisfying $\underline{\mathcal{N}} \subseteq S \subseteq \mathcal{N}$, let us see what further conditions must be imposed on S so that this solution also satisfies (KT-3) and (KT-4).

Consider (KT-4), which can be written

(KT-4) $$\lambda_{k_i} \geq 0, \qquad i = 1,\ldots,s.$$

In view of (6) and (10), (KT-4) is equivalent to

$$\frac{(\eta^1 + \eta^2)\eta^1_{i+1}}{(\eta^1_{i+1} + \eta^2_{i+1})} - \frac{(\eta^1 + \eta^2)\eta^1_i}{(\eta^1_i + \eta^2_{i+1})} \geq 0, \qquad i = 1,\ldots,s,$$

or

(12) $$(\eta_i^2/\eta_i^1) \geq (\eta^2_{i+1}/\eta^1_{i+1}), \qquad i = 1,\ldots,s.$$

Consider now (KT-3), which can be written

(KT-3A) $$\sum_{i=1}^{k} y_i^1 \geq \sum_{i} y_i^2 : x_i^2 < x^1_{k+1}, \quad k \notin S.$$

For the sake of simplicity, we have not bothered to exclude the redundant constraints among (2.1). In view of (KT-2), (KT-3A) amounts to requiring for $i = 1,\ldots,s+1$ and all $k \notin S$ but $k \in p_i$,

$$\left[\sum_{j\in p_i}^{k} y_j^1 : j < k\right] \ge \left[\sum_{j\in q_i} y_j^2 : x_j^2 < x_{k+1}^1\right].$$

Using (10), (11) and Lemma 3, this becomes

$$\left[\frac{(\eta_i^1 + \eta_i^2)}{\eta_i^1(\eta^1 + \eta^2)} \sum_{j\in p_i} h_j^1 : j \le k\right] \ge \left[\frac{(\eta_i^1 + \eta_i^2)}{\eta_i^2(\eta^1 + \eta^2)} \sum_{j\in q_i} h_j^2 : x_j^2 < x_{k+1}^1\right]$$

for $i = 1,\ldots,s+1$ and $k \in p_i$. For $i = 1,\ldots,s+1$, a little manipulation yields

$$(13)_i \qquad (\eta_i^2/\eta_i^1) \ge \left[\frac{\sum_{j\in q_i} h_j^2 : x_j^2 < x_{k+1}^1}{\sum_{j\in p_i} h_j^1 : j \le k}\right] \quad \text{for all } k \in p_i \text{ but } k \notin S.$$

We have proven, in view of Lemmas 3 and 4

__Lemma 5__: Let S satisfy $\underline{\mathcal{N}} \subseteq S \subseteq \overline{\mathcal{N}}$.

A necessary and sufficient condition on S in order that the unique solution of (KT-1) and (KT-2) also satisfy (KT-3) and (KT-4) is given by (12) and (13).

Lemmas 3, 4, and 5 immediately yield, in view of Theorem 2, the central result of this section.

__Theorem 3__: Let S satisfy $\underline{\mathcal{N}} \subseteq S \subseteq \overline{\mathcal{N}}$ and also (12) and (13). Then the optimal solution $\{\hat{y}_i^j\}$ of (P) is given by (10) and (11).

Brunk, et al. [2] has given a simple geometrical method for constructing S so as to fulfill the conditions of Theorem 2. Conditions (10) and (11) have purposely been written in such a form as to make this construction clear. It will not be repeated here, since it cannot be generalized to the case of primary interest to us.

3.2 THE CASE $N \geq 3$

When the partial differentiation in (KT-1) is carried out, the result is

$$(\text{KT-1})_i^1 \qquad (h_i^1/y_i^1) + \mu^1 + \sum_{k \in S^1} \left[\lambda_k^1 : k \geq i\right] = 0, \quad i = 1,\ldots,n_1,$$

$$(\text{KT-1})_i^j \quad (h_i^j/y_i^j) + \mu^j - \sum_{k \in S^{j-1}} \left[\lambda_k^{j-1} : x_i^j < x_{k+1}^{j-1}\right] + \sum_{k \in S^j} \left[\lambda_k^j : k \geq i\right] = 0,$$

$$i = 1,\ldots,n_j; \quad j = 2,\ldots,N-1,$$

$$(\text{KT-1})_i^N \qquad (h_i^j/y_i^j) + \mu^N - \sum_{k \in S^{N-1}} \left[\lambda_k^{N-1} : x_i^N < x_{k+1}^{N-1}\right] = 0, \quad i = 1,\ldots,n_N.$$

Let S^j be expressed as $\{k_1^j,\ldots,k_{s_j}^j\}$, where $k_1^j < \ldots < k_{s_j}^j$, when S^j is not empty. We see from (KT-1) that the ratios (h_i^j/y_i^j) remain constant over certain integer intervals. Toward identifying these intervals, we make the following definitions. Let P^j be a partition of $\{1,\ldots,n_j\}$ induced by S^j for $j = 1,\ldots,N-1$, as follows: $[1, k_1^j]$, $(k_1^j, k_2^j],\ldots,(k_{s_j}^j, n_j]$. Let Q^j be a partition of $\{1,\ldots,n_{j+1}\}$ induced by S^j for $j = 1,\ldots,N-1$, as follows (for better legibility, complicated subscripts are placed in parentheses):

$$\{i : x_i^{j+1} < x^j(k_1^j+1)\}, \{i : x^j(k_1^j+1) \leq x_i^{j+1} < x^j(k_2^j+1)\},\ldots,$$
$$\{i : x^j(k_{s_j-1}^j+1) \leq x_i^{j+1} < x^j(k_{s_j}^j+1)\}, \{i : x^j(k_{s_j}^j+1) \leq x_i^{j+1}\}.$$

For $j = 2,\ldots,N-1$, let $\Pi^j \equiv P^j \cdot Q^{j-1}$ be the product partition of $\{1,\ldots,n_j\}$ determined by P^j and Q^{j-1}. Let $\Pi^1 \equiv P^1$ and $\Pi^N \equiv Q^{N-1}$. We denote a generic cell of P^j, Q^j, or Π^j by a subscript, and define $\eta^j_p \equiv \sum_{i \in \Pi^j_p} h^j_i$ for each $p \in \Pi^j$ and $j = 1,\ldots,N$.

From (KT-1) we see that (h^j_i/y^j_i) remains constant within each cell of Π^j, $j = 1,\ldots,N$. Therefore, to determine $\{y^j_i\}$ satisfying (KT-1) it suffices to determine the ratios of (h^j_i/y^j_i) within these cells. We denote these ratios by $(h/y)^j_p$. The symbol $(y/h)^j_p$, used in the sequel, denotes the reciprocal. We have

<u>Lemma 3A</u>: If (KT-1) has a solution $\{y^j_i\}$, then $(h^j_i/y^j_i) = (h/y)^j_p$ for each i in cell p of Π^j, $j = 1,\ldots,N$.

We turn now to (KT-2). It can be written as (14) together with $(15)^j_p$ for each $p \in P^j$, $j = 1,\ldots,N-1$:

(14)
$$\sum_{i=1}^{n_1} y^1_i = 1,$$

$(15)^j_p$
$$\sum_{i \in P^j_p} y^j_i = \sum_{i \in Q^j_p} y^{j+1}_i .$$

By Lemma 3A, (14) may be rewritten as

(KT-2A)
$$\sum_{p \in P^1} \eta^1_p \, (y/h)^1_p = 1,$$

and $(15)^j_p$ as

$(\text{KT-2B})^1_p$
$$\eta^1_p (y/h)^1_p = \sum_{\pi \in P^2 \cdot Q^1_p} \eta^2_\pi (y/h)^2_\pi ,$$

$$(\text{KT-2B})^j_p \qquad \sum_{\pi \in P^j_p \cdot Q^{j-1}} \eta^j_\pi (y/h)^j_\pi = \sum_{\pi \in P^{j+1} \cdot Q^j_p} \eta^{j+1}_\pi (y/h)^{j+1}_\pi ,$$

for $j = 2,\ldots,N-2$ (for $N \geq 4$), and

$$(\text{KT-2B})^{N-1}_p \qquad \sum_{\pi \in P^{N-1}_p \cdot Q^{N-2}} \eta^{N-1}_\pi (y/h)^{N-1}_\pi = \eta^N_p (y/h)^N_p .$$

This assumes, of course, that all $h^j_i > 0$, since (y^j_i/h^j_i) is undefined otherwise. When $h^j_i = 0$ (this can happen only for $i = 1$ or $i = n_j$), the corresponding term in (14) or (15) should replace the offending term in (KT-2A) or $(\text{KT-2B})^j_p$; that is, y^j_i replaces* the corresponding $\eta^j_\pi (y/h)^j_\pi$.

Let us see what can be done to solve (KT-1) and (KT-2) for the μ^j, λ^j_k, and y^j_i. From $(\text{KT-1})^j_{n_j}$ we see that

$$(16) \qquad \mu^j = -(h/y)^j_{n_j} , \qquad j = 1,\ldots,N.$$

Upon subtracting $(\text{KT-1})^1_{k^1_i+1}$ from $(\text{KT-1})^1_{k^1_i}$, one obtains for $i = 1,\ldots,s_1$,

$$(17)^1 \qquad \lambda^1_{k^1_i} = (h^1_{k^1_i+1}/y^1_{k^1_i+1}) - (h^1_{k^1_i}/y^1_{k^1_i}).$$

One similarly obtains for $j = 2,\ldots,N-1$ and $i = 1,\ldots,s_j$,

---

*It is helpful to recall that our assumption $\underline{\pi}^j \subseteq S^j \subseteq \pi^j$ ($j = 1,\ldots,N-1$) implies that when $h^j_i = 0$, we have $\eta^j_\pi = 0$ for the cell Π^j_π containing i.

$$(17)^j \qquad \lambda^j_{k^j_i} = (h^j_{k^j_i+1}/y^j_{k^j_i+1}) - (h^j_{k^j_i}/y^j_{k^j_i})$$

$$+ \sum_{k \in S^{j-1}} \lambda^{j-1}_k : x^j_{k^j_i} < x^{j-1}_{k+1} \le x^j_{k^j_i+1} .$$

Thus one can express the μ^j and λ^j_k in terms of the y^j_i.

By an appropriate linear transformation of the system (KT-1), one can eliminate the μ^j and λ^j_k. The resulting system, taken with (KT-2A) and (KT-2B), is nonlinear and does not yield to analytical solution techniques. This is in contrast to the case N = 2, and destroys our hopes of obtaining generalizations of Lemmas 4 and 5 for $N \geq 3$. Lemma 3A and (17) will be of considerable value in what follows, however, as will the following weaker version of Lemma 4.

<u>Lemma 4A</u>: Let S^j satisfy $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$, $j = 1,\ldots,N-1$. Then (KT-1) and (KT-2) have a unique solution $(y^j_i;\ \lambda^j_k;\ \mu^j)$. Moreover, all $y^j_i > 0$ in this solution, and the μ^j and λ^j_k are given in terms of the y^j_k by relations (16) and (17), respectively.

<u>Proof</u>: Consider the following subproblem derived from (P) according to $S^j (j=1,\ldots,N-1)$, where $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$:

$$(P_S) \qquad \underset{y^j_i \geq 0}{\text{Maximize}} \sum_{j=1}^{N} \sum_{i=1}^{n_j} h^j_i \ln y^j_i \text{ subject to (KT-2).}$$

The proof of Theorem 1 is easily adapted to show that (P_S) has a unique optimal solution, and that this solution has all positive components. Applying the Kuhn-Tucker Theorem to (P_S) as in Theorem 2, we see that (KT-1) and (KT-2) are in fact the Kuhn-Tucker conditions associated with (P_S). Hence (KT-1) and (KT-2) must have a solution, and $\{y_i^j\}$ is unique and positive in any such solution. The above manipulations show that the μ^j and λ_k^j must also be unique, since they can be expressed in terms of the y_i^j according to (16) and (17). This completes the proof.

IV. A REDUCTION PROCEDURE

The basic strategy of the last section was to determine, with the help of a general solution of (KT-1) and (KT-2), conditions that the S^j must satisfy (in addition to $\underline{\mathcal{N}}^j \subseteq S^j \subseteq \mathcal{N}^j$) so that (KT-3) and (KT-4) will have a solution. With the S^j so determined, the optimal solution of (P) is obtained from the corresponding solution of (KT-1) and (KT-2). This strategy was completely successful only in the case $N = 2$, the main obstacle for the case $N \geq 3$ being the apparent impossibility of obtaining a general solution of (KT-1) and (KT-2). In this section we essentially pursue the same strategy, but in an iterative numerical manner. Lemmas 3A and 4A will be quite valuable in this pursuit, and the result is an efficient computational procedure.

The proof of Lemma 4A indicates that solving the concave program

$$(P_S) \qquad \underset{y_i^j > 0}{\text{Maximize}} \sum_{j=1}^{N} \sum_{i=1}^{n_j} h_i^j \ln y_i^j \qquad \text{subject to (KT-2)}$$

is equivalent to solving (KT-1,2). Our strategy then is to choose an initial S and solve (P_S), and determine whether the resulting $(y_i^j; \mu^j; \lambda_k^j)$ satisfies (KT-3) and (KT-4); if so, then the current solution solves (P) and we may terminate, but if not, then S is modified according to the sign reversals found in (KT-3) and (KT-4), and the new (P_S) is solved, and so on. After a usually finite number of iterations, a subproblem yielding the optimal solution of (P) is

found. A generalization of this strategy has been presented by the author elsewhere [4, 5]. When the results therein are specialized in view of Lemma 4A to the present problem, the following reduction procedure and theorem are obtained.

Reduction Procedure

Step 1: Choose an initial S to satisfy $\underline{\eta}^j \subseteq S^j \subseteq \overline{\eta}^j$, $j = 1,\ldots,N-1$.

Step 2: Solve (P_S) for its unique optimal solution $\{y_i^j\}$. Compute $\{\lambda_k^j\}$ by (17). Check (KT-3) and (KT-4): if both are satisfied, then terminate $\left(\{y_i^j\}\text{ solves }(P)\right)$; otherwise, let $T^j (j = 1,\ldots,N-1)$ be the indices of the sign reversals found in (KT-3,4), that is, let T^j be

$$\{k \in \overline{\eta}^j - S^j : \sum_{i=1}^{k} y_i^j < \sum_i y_i^{j+1} : x_i^{j+1} < x_{k+1}^j\} \cup \{k \in S^j - \underline{\eta}^j : \lambda_k^j < 0\},$$

and go to Step 3.

Step 3: Pick t at random from $\underset{j=1}{\overset{N-1}{\times}} T^j$. Replace* S^j by $S^j \pm t$, where j is the index satisfying $t \in T^j$. Return to Step 1.

Theorem 4: The reduction procedure is well defined, and terminates in a finite number of steps with probability 1.

The computational efficiency of the procedure depends on (i) the time required to solve a typical subproblem, and (ii) the number of subproblems that must be solved before termination. We consider each of these factors in turn.

---

*$S^j \pm t$ means $S^j - t$ if $t \in S^j$, and $S^j \cup t$ if $t \notin S^j$.

SOLVING A SUBPROBLEM

The subproblems have the same objective function as (P), but fewer constraints, and these are equalities rather than inequalities. Moreover, the result of Lemma 3A makes it a trivial matter to substantially reduce the number of variables (especially when S is sparse); all that is needed is one variable for each cell of the partitions $\pi^j(j=1,\ldots,N)$ induced by S. When this is done, the constraints (KT-2) can be rewritten as (KT-2A) and (KT-2B), and the objective function (after a simple manipulation) as

$$\sum_{j=1}^{N} \sum_{p \in \pi^j} \pi_p^j \ln (y/h)_p^j + \text{a constant}, \tag{18}$$

where it is understood that if $h_i^j = 0$, then the term $\pi_p^j \ln (y/h)_p^j$ is omitted for the cell π_p^j containing i (cf. the next to last footnote). Hence (P_S), thus simplified, is considerably more convenient to solve than (P) itself.

Even further simplifications are possible, in fact, by solving the constraints (KT-2A) and (KT-2B) explicitly for (s_j+1) of the $(y/h)_p^{j+1}$ $(j=1,\ldots,N-1)$ and one of the $(y/h)_p^1$. Because of the convenient triangular structure of the constraints this can be done easily without performing a matrix inversion. If the expressions for the dependent variables are substituted into the objective function, only the nonnegativity constraints are left in (P_S). And even these can be ignored if (P_S) is to be solved by an optimal gradient-type

procedure, for $\ln (y/h)^j_p \to -\infty$ as $(y/h)^j_p \to 0$ implies that boundary-repulsion occurs automatically.*

All this is to say that (P_S) can be rewritten as an essentially unconstrained problem in many fewer variables than (P). Thus the time required to solve a typical subproblem can be expected to be very small compared with that necessary to solve (P) itself by a comparable gradient method.

THE NUMBER OF SUBPROBLEMS

It seems difficult to derive theoretical results concerning the number of subproblems that must be solved before termination. The Markov chain analysis and computational results reported by this writer in [5], however, suggest the following tentative hypothesis: the number of required subproblems can be expected to be about twice the number of mistakes made in identifying, through the choice of the initial subproblem, the constraints of (P) that are actually binding at the optimal solution. This hypothesis holds provided that the sign reversals of (KT-3) and (KT-4) are a sufficiently reliable guide regarding which constraints are actually binding at the optimum. If the hypothesis is even approximately valid, then use of the reduction procedure is greatly encouraged; for it has been our experience with numerous test problems (cf. the next section) that usually only a

---

*We assume that the gradient procedure is initiated at a positive feasible solution of (P_S), which can always be easily done (cf. the proof of Theorem 1). Strictly speaking, boundary-repulsion only occurs directly for the ratios actually in (18); it occurs indirectly for the y^j_i such that $h^j_i = 0$, however, because $\underline{\mathcal{H}}^j \subseteq S^j$.

comparatively small proportion of all constraints are actually binding at the optimal solution, and hence by taking $\underline{\mathcal{T}}^j$ as the initial S^j, one would be assured that the number of subproblems to be solved is small.

We turn now to the computational studies of these matters.

V. COMPUTATIONAL EXPERIENCE

IMPLEMENTATION

Preliminary tests designed to study the tentative hypothesis mentioned in the previous section quickly confirmed that the sign reversals found in (KT-3) and (KT-4) at Step 2 are indeed a remarkably reliable guide in identifying the constraints actually binding at the optimum. To take fuller advantage of this situation, it was decided to implement a slight variant of the reduction procedure, namely: at the first few (five) executions of Step 3 utilize all sign reversals recorded in T^j to modify each S^j, instead of using just one at random. Theorem 4 obviously remains valid for this variant.

In considering how to solve the subproblems (P_S), we decided to take a major shortcut: the (triangular) constraints of (P_S) were not used to solve for and eliminate some of the variables. Undoubtedly, the price of this expediency is significantly longer solution times for the subproblems, but it was adopted for the sake of programming simplicity since the primary object of the investigation was the number of subproblems required by the reduction procedure.

We chose a general nonlinear programming routine due to Graves [6] for use in the subroutine for solving (P_S). This is a sophisticated large-step gradient method that has proven quite effective in a variety of applications. An important issue regarding its use concerns the choice of an appropriate termination criterion. Fortunately, the routine computes an upper bound $\bar{v}$ on the true optimal value of (P_S) at each iteration, and so a bound on the maximum possible error is always available: $\bar{v}$ minus the value of the current feasible solution.

We therefore provided a parameter TL ("termination level") to cause return from Graves' routine when the maximum possible relative error for (P_S) drops below TL $\times$ 100 percent. For large problems, of course, it may not be possible to reach the desired value of TL in a reasonable time. Thus a parameter ICYCLE is also used to cause a return from Graves' routine after ICYCLE iterations.

S P. Azen carried out the implementation in Fortran IV for the IBM 7044. Details of the test program, called ESOD-3, are available in [1].

TEST PROBLEMS

ESOD-3 has been tested extensively on dozens of different problems. All have been solved successfully. We present in Table 1 only our experience with some of the larger problems, the largest of these having 381 variables and 123 nonredundant constraints (not counting sign restrictions on the variables).

Series A consists of three distributions ($N = 3$), all exponential, with means 1, 2 and 3, respectively. Problem A1 has 15, 12 and 9 observations, respectively, from each distribution. When the conversion is made from the constrained maximum liklihood estimation problem to a concave program of the form (P), A1 has 37 variables and 13 non-redundant constraints. The number of observations from the three distributions increases in multiples of 5, 4, and 3, respectively, in problems A2 through A6.

Series B consists of 4 Weibull distributions with scale parameter λ and shape parameter α as follows:

Distribution	λ	α	Mean
1	5	1	0.2
2	1	2	0.88263
3	1	3	0.89338
4	0.1	1	10.0

Problems B1 through B4 each have 10, 20, 30, and 38 observations from each distribution, respectively.

Series C consists of 10 normal distributions with unit variance and means 0, 2, 4, 6, 7, 8, 9, 9½, 10, 10½, respectively. Problems C1 through C4 also have 10, 20, 30, and 38 observations from each distribution.

Table 1 summarizes our computational experience with these problems. All were run with TL = 0.02, and with the initial $s^j = \underline{\eta}^j$. Series A was run with ICYCLE = 15, series B with ICYCLE = 25, and series C with ICYCLE = 30. These values of ICYCLE are quite small, and cause the subproblems to be considerably suboptimized; this was done deliberately to take advantage of the empirical observation that the ultimate sign reversals in (KT-3) and (KT-4) manifest themselves long before (P_S) is solved optimally. The parameter ICYCLE is automatically made inoperative in the final subproblem, so that TL becomes effective as the termination criterion. In some of the larger problems, it will be noted that the 2-percent level was not achieved; in these cases the permissible step length SK in Graves' routine became 0, so that no further improvement was numerically possible.

Table 1

SOME COMPUTATIONAL EXPERIENCE WITH THE ESOD-3 PROGRAM

Problem	No. of Variables $\sum_{j=1}^{N} n_j$	No. of Constraints (Nonredundant)	No. of Subproblems Required	Solution Time (7044)	Guaranteed Maximum Possible Error (%)
A1	37	13	3	18"	1.85
A2	48	15	3	17"	1.33
A3	60	17	4	23"	1.30
A4	72	20	4	22"	1.93
A5	84	27	3	23"	1.33
A6	96	30	5	39"	2.56
B1	43	9	1	4"	1.73
B2	83	23	3	53"	1.89
B3	122	33	2	49"	6.85
B4	154	39	2	1'45"	6.90
C1	103	26	2	2'20"	1.82
C2	203	71	3	4'25"	2.53
C3	302	96	4	5'06"	5.69
C4	381	123	3	4'27"	6.85

CONCLUSIONS

The central conclusion from this experience is that the number of subproblems that must be solved remains extremely small, even when the problem becomes quite large. This remarkable result is certainly a stronger justification for using the present reduction strategy than could reasonably have been hoped for in advance.

This experience also shows that, in spite of the efficiency-reducing shortcut mentioned above under implementation, ESOD-3 can solve large practical problems in a reasonable time. The guaranteed maximum possible error levels attained for the optimal value of (P) were generally sufficient to assure three- or four-place accuracy in the estimated frequencies--probably more than sufficient for most uses.

It is of interest to compare these computing times and accuracies with results obtained by applying Graves' routine directly to (P) itself (without using the reduction procedure). For problem A1, in 8 minutes and over 300 iterations the guaranteed maximum possible error (GMPE) was 30 percent; in 20 minutes and over 1000 iterations the GMPE was 18 percent. For problem A2, the GMPE was 600 percent in 8 minutes and 34 percent in 30 minutes. The performance on problems A3 through A6 was even poorer when compared with the results shown in Table 1. The story was similar for Series B and C.* This reflects the usual tendency of first-order gradient methods to converge slowly in the neighborhood of the optimum.** That the very same gradient method should perform so much better in terms of time and accuracy, when used in conjunction with the present reduction strategy, constitutes a compelling example of the power of this strategy.

---

*Some of the larger problems could not even be attempted, as the direct application of Graves' routine to (P) is less economical in terms of primary computer storage than the present approach.

**(P) and (P_S) appear to be particularly difficult and ill-behaved in this regard.

REFERENCES

1. Azen, S. P., Description of the Program ESOD-3 for Calculating Constrained Maximum Liklihood Estimates of N Stochastically Ordered Distributions, The RAND Corporation, RM-5536-PR, July 1968.

2. Brunk, H. D., et al., "Maximum Liklihood Estimation of the Distributions of Two Stochastically Ordered Random Variables," Journal of the American Statistical Association, Vol. 61, No. 316, December 1966, pp. 1067-1080.

3. Kuhn, H., and A. Tucker, "Nonlinear Programming," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, 1951, pp. 481-492.

4. Geoffrion, A. M., "Reducing Concave Programs with Some Linear Constraints," SIAM Journal on Applied Mathematics, Vol. 15, No. 3, May 1967, pp. 653-664.

5. ---------, "A Markovian Procedure for Strictly Concave Programming with Some Linear Constraints," to appear in Proceedings of the Fourth International Conference on Operational Research, Cambridge, Massachusetts, 1966.

6. Graves, G. W., "A Non-Linear Programming Procedure," to appear as a Western Management Science Institute Working Paper.

7. Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming: Part I, Linear Constraints," SIAM Journal, Vol. 8, pp. 181-217.

8. Miller, C. E., "The Simplex Method for Local Separable Programming," in R. L. Graves and P. Wolfe (eds.), Recent Advances in Mathematical Programming, McGraw-Hill, New York, 1963, pp. 89-100.

9. Scheuer, E. M., R. E. Barlow and F. Proschan, Maximum Likelihood Estimation and Conservative Confidence Interval Procedures in Reliability Growth and Debugging Problems, The RAND Corporation, RM-4749-NASA, January 1966.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)
Western Management Science Institute

2a REPORT SECURITY CLASSIFICATION
Unclassified

2b GROUP

3 REPORT TITLE
Constrained Maximum Likelihood Estimation of N Stochastically Ordered Distribution.

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

5 AUTHOR(S) (Last name, first name, initial)
Geoffrion, Arthur M.

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
July, 1968	34	9

8a CONTRACT OR GRANT NO
Nonr 233(75)

b PROJECT NO

c

d

9a ORIGINATOR'S REPORT NUMBER(S)

9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited.

Western Management Science Institute
University of California
Los Angeles, California 90024

11 SUPPLEMENTARY NOTES

12 SPONSORING MILITARY ACTIVITY

13 ABSTRACT

See Summary

DD FORM 1 JAN 64 1473 0101-807-6800

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
maximum likelihood estimation mathematical programming large-scale optimization stochastic ordering statistics						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

Security Classification